Genetic Epidemiology 14:827–832 (1997)

Sib-pair Linkage Analyses of Nuclear Family Data: Quantitative Versus Dichotomous Disease Classification

Jeannette F. Korczak and Alisa M. Goldstein

Genetic Epidemiology Branch, National Cancer Institute, Bethesda, Maryland

Model-free sib-pair linkage analysis was used to screen 367 highly polymorphic markers for evidence of linkage to a disease, defined either quantitatively (Q1) or dichotomously (AF). Five individual replicates, plus a case family data set containing all families in these replicates with at least one individual with AF, were analyzed. Sib-pair linkage results for Q1 and AF varied considerably among the five replicates and did not consistently detect any of the three underlying major loci, MG1, MG2, and MG3. For the pooled case families, linkage analyses of Q1, but not AF, detected the flanking markers for MG1 and MG2 at the 0.05 and 0.01 levels, respectively. Overall, type 1 error rates were not elevated. The ability to analyze the disease quantitatively (Q1) and construct a data set more appropriate for linkage analysis (case families) enhanced the power to detect at least some of the major loci underlying the disease. © 1997 Wiley-Liss, Inc.

Key words: complex traits, genome scanning, nonparametric linkage analysis, oligogenic traits

INTRODUCTION

Large simulated data sets provide an opportunity to examine the methods currently widely used to localize disease loci. Previously (GAW9, Problem 1), we used model-free sib-pair linkage analysis for a single replicate of 200 nuclear case families with a dichotomous disease and genotypes at 360 highly polymorphic markers and were unable to identify any of the disease loci [Korczak et al., 1995]. All highly significant marker loci that were identified proved to be false indications of linkage but did not suggest an elevated type 1 error rate. These findings were not entirely unexpected, given the particular simulation and the limited number of multiplex families in these data. The GAW10 nuclear family data set (Problem 2A) provided another opportunity to explore the

Address reprint requests to Dr. Jeannette F. Korczak, Genetic Epidemiology Branch, NCI, Executive Plaza North, Room 400, 6130 Executive Boulevard MSC7360, Bethesda, MD 20892-7360.

power and type 1 error rate of model-free sib-pair linkage analysis. Specifically, our goals were to (i) identify loci underlying the disease phenotype, expressed either quantitatively (Q1) or dichotomously (AF), with or without the inclusion of other relevant traits as covariates and (ii) determine the consistency of the linkage results obtained from different replicates of the data.

MATERIALS AND METHODS

We selected at random five of the 200 replicates of the Problem 2A data for model-free sib-pair linkage analysis: Replicates 29, 46, 80, 169, and 180. These data are described in detail in another paper in this volume. Briefly, each replicate consisted of 239 randomly ascertained nuclear families with 1,164 individuals (1,000 living) of whom about 7% were classified as having AF after imposing a threshold on Q1. The majority of families in each replicate contained no AF individuals (control families). In our sample, the number of control families ranged from 169-193 per replicate; the numbers of simplex and multiplex case families varied from 38-56 and 8-14 per replicate, respectively. Because of the small number of case families and concordantly affected sib- pairs in individual replicates (5-20 pairs/replicate), all simplex and multiplex families from the five replicates were combined to form a pooled case family data set with 293 families of which 55 had two to five affected members. This pooled data set contained 1,275 living individuals, of whom 368 (~29%) had AF.

From the living parents in the five replicates, as well as those in the pooled case families, we obtained Pearson correlation coefficients for Q1-Q5, EF, and age and performed backward stepwise regression for Q1 and AF to determine the most significant of Q2-Q5, EF, age, and sex for inclusion as covariates in model-free sib-pair linkage analyses of Q1 and AF. We also obtained familial correlations for Q1 from the data on all living individuals using the program FCOR [S.A.G.E, 1994].

For the individual replicates and the pooled case families, model-free sib-pair linkage analyses [Haseman and Elston, 1972; Elston et al., 1973; Elston, 1984] as implemented in the program SIBPAL [S.A.G.E., 1994] were performed to screen the 367 highly polymorphic markers for evidence of linkage to Q1 (untransformed and under square root and natural log transformations) and AF. SIBPAL regresses the squared sib-pair trait difference on the estimated proportion of alleles identical by descent ($\hat{\pi}$) at each marker locus. Marker genotypes for all individuals, alive and deceased, were included in the analyses. Since the markers were codominant and all individuals were typed, the actual marker allele frequencies were not needed to determine the proportion of alleles identical by descent. Thus for convenience we assumed each marker locus had 15 equally frequent alleles. For deceased individuals, Q1-Q5, AF, and EF were coded as missing.

RESULTS

Analyses of Five Individual Replicates

Pearson correlations of Q1 with Q2, Q3, EF, and age were significant at the 0.001 level in each replicate, whereas those with Q4 and Q5 were variable and often nonsignificant. Correlations of Q1 with Q3 were strongest, averaging 0.63 over the five replicates (N = 314 living parents per replicate). In four replicates the parent-offspring

and sibling correlations were of similar magnitude ($r \sim 0.2$) and approximately 3 times greater than the spouse correlation, suggestive that an underlying genetic component accounts for at least some of the variation in Q1.

In all sib-pair linkage analyses of Q1, the results using untransformed data were very similar to those from either natural log or square root transformations. Thus only the untransformed results are reported here. Table I shows p-values for marker loci flanking MG1 (D5G14 and D5G15), MG2 (D8G26 and D8G27), and MG3 (D4G14 and D4G15), plus all markers with negative regressions nominally significant at the $\alpha = 0.01$ level for either Q1 or AF in at least one of five replicates. No covariates were included in these analyses. Table I reveals a lack of consistency in the sib-pair linkage results, especially among replicates for the same trait and, to a lesser extent, within replicates for Q1 versus AF. Of 34 markers with at least one highly significant negative regression, only D5G9 and D8G22 were significant at the $\alpha = 0.01$ level in more than one of the 10 analyses. The ability to detect MG1, MG2, or MG3 was very limited. Only D5G14 (MG1), D8G26 (MG2), and D8G27 (MG2) were detectable at the $\alpha = 0.01$ level, but each attained significance in only one of 10 analyses. The existence of an effect of MG3 on the disease (Q1 or AF) was not readily discernable.

To assess the type 1 error rate, keeping in mind that p-values for tightly linked markers are correlated, we first grouped together adjacent marker loci that were each nominally significant at the $\alpha = 0.01$ level into a "cluster," as previously described [Korczak et al., 1995]. After excluding those clusters that contained any of the markers flanking MG1, MG2, or MG3, the number of remaining clusters (i.e., type 1 errors) varied from one to seven in each of the 10 analyses and averaged 2.8 over these analyses. This is less than the three to four type 1 errors we would have expected at the $\alpha = 0.01$ level simply by chance, leading us to conclude that the type 1 error rate was not elevated.

Analyses of the Pooled Case Family Data

Pearson correlations of Q1 with Q2, Q3, EF, and age were significant at the 0.001 level, just as in the individual replicates, with the strongest correlation again between Q1 and Q3, $r = 0.73$ ($N = 383$ living parents). Backward stepwise regression for Q1 and AF retained Q2, Q3, and age for use as covariates in sib-pair linkage analyses.

Sib-pair linkage results for the untransformed and natural log and square root transformed Q1 values were very similar; only those for the untransformed trait are presented. Table II shows p-values for marker loci flanking MG1-MG3 or with negative regressions nominally significant at the $\alpha = 0.01$ level for either Q1 or AF in the pooled case families, with or without the joint inclusion of Q2, Q3, and age as covariates. Controlling for these covariate phenotypes had virtually no effect on the results of sib-pair linkage analysis for either Q1 or AF. Results for Q1 and AF were somewhat more consistent with each other than in the analyses of individual replicates, with p-values for AF usually less significant than those for Q1. Only in the analyses for Q1 was it possible to detect any of the underlying major loci: p-values for the markers flanking MG1 (D5G14 and D5G15) and MG2 (D8G26 and D8G27) were approximately 0.05 and 0.01, respectively. As was the case in the individual replicates, it was not possible to detect MG3. Based on clusters of adjacent markers, as previously defined, we observed on average 4.5 and 1.5 type 1 errors per analysis for Q1 and AF, respectively, not substantially greater than what was expected by chance alone.

TABLE I. p-Values for Marker Loci Flanking MG1, MG2, or MG3 or with Highly Significant ($\alpha = 0.01$) Negative Regressions of the Squared Sib-pair Trait Difference on $\hat{\pi}$ for Either Q1 or AF in at Least One of Five Replicates (444 Effective df)

	Trait Q1					Trait AF				
Replicate:	R29	R46	R80	R169	R180	R29	R46	R80	R169	R180
Markers:										
D1G18	0.81	0.33	0.68	0.40	0.17	0.54	0.83	0.96	0.73	0.003
D2G45	0.13	0.29	0.52	0.27	0.31	0.03	0.009	0.82	0.02	0.42
D4G5	0.83	0.19	0.42	0.07	0.96	0.42	0.70	0.19	0.0011	0.20
D4G14 [a]	0.77	0.06	0.08	0.76	0.98	0.67	0.47	0.92	0.26	0.89
D4G15	0.84	0.18	0.34	0.52	0.97	0.47	0.32	0.95	0.24	0.92
D4G29	0.62	0.004	0.42	0.16	0.86	0.15	0.04	0.15	0.22	0.84
D5G2	0.18	0.74	0.91	0.43	0.07	0.006	0.91	0.79	0.81	0.26
D5G9	0.06	0.007	0.22	0.009	0.09	0.07	0.58	0.58	0.07	0.22
D5G10	0.010	0.16	0.62	0.33	0.08	0.02	0.73	0.12	0.34	0.34
D5G11	0.14	0.02	0.27	0.23	0.005	0.12	0.84	0.36	0.68	0.42
D5G12	0.12	0.22	0.49	0.59	0.12	0.009	0.63	0.29	0.77	0.54
D5G14	0.06	0.08	0.24	0.31	0.09	0.007	0.82	0.21	0.55	0.42
D5G15	0.03	0.18	0.14	0.27	0.06	0.26	0.70	0.37	0.56	0.63
D6G6	0.28	0.92	0.29	0.25	0.38	0.42	0.51	0.20	0.005	0.21
D6G44	0.69	0.29	0.35	0.87	0.88	0.40	0.006	0.33	0.76	0.99
D6G45	0.83	0.26	0.15	0.84	0.95	0.55	0.008	0.48	0.47	0.70
D6G46	0.78	0.31	0.06	0.69	0.96	0.39	0.007	0.75	0.46	0.95
D8G9	0.20	0.005	0.10	0.47	0.18	0.14	0.20	0.14	0.64	0.39
D8G11	0.12	0.006	0.02	0.50	0.02	0.08	0.27	0.29	0.62	0.64
D8G17	0.29	0.01	0.03	0.06	0.08	0.005	0.79	0.27	0.06	0.85
D8G21	0.47	0.26	0.19	0.15	0.004	0.10	0.18	0.13	0.17	0.87
D8G22	0.68	0.50	0.004	0.57	0.12	0.28	0.41	0.008	0.24	0.97
D8G26	0.11	0.30	0.012	0.06	0.008	0.25	0.67	0.07	0.69	0.96
D8G27	0.33	0.52	0.06	0.10	0.002	0.89	0.92	0.18	0.26	0.55
D9G3	0.62	0.54	0.68	0.42	0.49	0.93	0.005	0.38	0.34	0.98
D9G5	0.98	0.66	0.42	0.79	0.73	0.90	0.002	0.45	0.52	0.93
D9G6	0.35	0.83	0.43	0.80	0.45	0.94	0.008	0.56	0.12	0.93
D9G7	0.50	0.34	0.68	0.53	0.35	0.84	0.002	0.42	0.45	0.88
D9G8	0.32	0.26	0.44	0.42	0.22	0.70	0.003	0.09	0.40	0.91
D9G9	0.44	0.60	0.80	0.29	0.30	0.56	0.002	0.44	0.48	0.78
D9G16	0.004	0.68	0.24	0.82	0.12	0.43	0.51	0.61	0.63	0.80
D9G35	0.67	0.05	0.80	0.17	0.56	0.86	0.007	0.86	0.62	0.44
D9G36	0.45	0.007	0.38	0.15	0.40	0.90	0.03	0.85	0.64	0.28
D10G13	0.42	0.006	0.11	0.61	0.84	0.99	0.06	0.36	0.28	0.18
D10G20	0.34	0.054	0.32	0.40	0.54	0.67	0.002	0.09	0.38	0.51
D10G21	0.26	0.15	0.0004	0.56	0.31	0.62	0.08	0.04	0.49	0.08
D10G29	0.02	0.51	0.10	0.79	0.38	0.009	0.05	0.31	0.88	0.16

[a] Bold-faced markers on chromosomes 4, 5, and 8 flank MG3, MG1, and MG2, respectively.

DISCUSSION

Sib-pair linkage results from the individual replicates demonstrated an overall lack of power to detect any of the three major loci accounting for a portion of the variation in the disease, either expressed quantitatively (Q1) or dichotomously (AF). From information provided to GAW10 participants, these loci accounted for approximately 40% of the variance of Q1, including the additive x additive epistatic interaction between MG1 and MG2. Although the specific clusters of marker loci with highly significant negative

TABLE II. p-Values for Marker Loci Flanking MG1, MG2, or MG3 or with Highly Significant (α = 0.01) Negative Regressions of the Squared Sib-pair Trait Difference on $\hat{\pi}$ for Either Q1 or AF in the Pooled Case Families, With or Without Covariates (596 Effective df)

	Trait Q1		Trait AF	
Covariates:	None	Q2, Q3, Age	None	Q2, Q3, Age
Markers:				
D1G37	0.007	0.014	0.08	0.11
D1G38	0.02	0.009	0.015	0.012
D4G14 [a]	0.42	0.67	0.58	0.68
D4G15	0.17	0.48	0.37	0.54
D4G29	0.014	0.08	0.002	0.007
D5G8	0.008	0.006	0.35	0.40
D5G9	0.001	0.004	0.12	0.19
D5G14	0.08	0.04	0.14	0.13
D5G15	0.14	0.06	0.55	0.51
D5G17	0.03	0.002	0.32	0.22
D8G10	0.008	0.008	0.054	0.07
D8G11	0.0101	0.002	0.23	0.19
D8G14	0.007	0.014	0.29	0.38
D8G15	0.001	0.004	0.09	0.15
D8G16	0.001	0.003	0.14	0.20
D8G17	0.004	0.004	0.13	0.15
D8G26	0.008	0.006	0.60	0.64
D8G27	0.012	0.011	0.79	0.82
D8G28	0.006	0.008	0.42	0.48
D8G29	0.007	0.004	0.46	0.47
D10G20	0.21	0.28	0.008	0.0104

[a] Bold-faced markers on chromosomes 4, 5, and 8 flank MG3, MG1, and MG2, respectively.

regressions varied considerably among replicates, the number of such clusters was not great enough to suggest an elevated type 1 error rate.

The results for the pooled case families should be interpreted as preliminary, since they are based on a single replicate of data. Sib-pair linkage analyses of Q1 showed greater power to detect two of three underlying major loci than was seen in the individual replicates. None of these loci were detected in sib-pair analyses of the dichotomous trait, AF. This is most likely the consequence of loss of information inherent in the dichotomization process. Inclusion of significant covariates did not measurably increase the power of sib-pair linkage analysis for either trait in these data. The type 1 error rate did not appear to be significantly elevated.

These results, particularly for the dichotomous trait in individual replicates, are similar to those we reported at GAW9. There, too, we observed a lack of power but not a substantial increase in the type 1 error rate in sib-pair linkage analyses of a dichotomous disease, using a data set that also was not ideally suited to linkage analysis. For GAW10, the ability to analyze the disease not only dichotomously but also quantitatively and to construct a data set more appropriate for linkage analysis increased the power to detect at least some of the major genes underlying the disease through model-free sib-pair linkage analysis.

ACKNOWLEDGMENTS

The authors express their appreciation to Drs. Joan E. Bailey-Wilson, Karen A. Weissbecker, and Alexander F. Wilson for sharing insights and strategies for analyzing these data. The results of this paper were obtained by using the program package S.A.G.E., which is supported by a U.S. Public Health Service Resource grant RR03655 from the National Center for Research Resources.

REFERENCES

Elston RC (1984): Genetic Analysis Workshop II: Sib pair screening tests for linkage. Genet Epidemiol 1:175-178.

Elston RC, Kringlen E, Namboodiri KK (1973): Possible linkage relationships between certain blood groups and schizophrenia or other psychoses. Behav Genet 3:101-106.

Haseman JK, Elston RC (1972): The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3-19.

Korczak JF, Pugh EW, Premkumar S, Guo X, Elston RC, Bailey-Wilson JE (1995): Effects of marker information on sib-pair linkage analysis of a rare disease. Genet Epidemiol 12:625-630.

S.A.G.E. (1994): Statistical Analysis for Genetic Epidemiology, Release 2.2. Computer program package available from the Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland.